BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

STUDIES IN MICROBIOLOGY VOLUME 2

Principles of Gene Manipulation

AN INTRODUCTION TO GENETIC ENGINEERING

R. W. OLD MA, PhD
Department of Biological Sciences,
University of Warwick,
Coventry

S. B. PRIMROSE BSc, PhD
Searle Research and Development,
High Wycombe

THIRD EDITION

BLACKWELL SCIENTIFIC PUBLICATIONS
OXFORD LONDON EDINBURGH
BOSTON PALO ALTO MELBOURNE

Editorial offices:
Osney Mead, Oxford, OX2 0EL
8 John Street, London, WC1N 2ES
23 Ainslie Place, Edinburgh, EH3 6AJ
3 Cambridge Center, Suite 208, Cambridge, Massachusetts 02142, USA
667 Lytton Avenue, Palo Alto California 94301, USA
107 Barry Street, Carlton Victoria 3053, Australia

First published 1980
Second edition 1981
Reprinted 1982, 1983 (twice)
Third edition 1985
Reprinted 1986 (twice), 1987, 1988

Spanish edition 1986
Japanese edition 1987

Typeset by Acorn Bookwork, Salisbury, Wiltshire
Printed and bound in Great Britain at the University Press, Cambridge

DISTRIBUTORS

USA
Year Book Medical Publishers
200 North LaSalle Street
Chicago, Illinois 60601

Canada
The C.V. Mosby Company
5240 Finch Avenue East,
Scarborough, Ontario

Australia
Blackwell Scientific Publications (Australia) Pty Ltd
107 Barry Street
Carlton, Victoria 3053

British Library
Cataloguing in Publication Data

Old, R. W.
Principles of gene manipulation.—3rd ed.—
(Studies in microbiology; v.2)
1. Genetic engineering
I. Title II. Primrose, S. B.
III. Series
575.1 QH442

ISBN 0-632-01318-4

Chapter 1 Basic Techniques

Introduction

Occasionally technical developments in science occur that enable leaps forward in our knowledge and increase the potential for innovation. Molecular biology and biomedical research have recently experienced just such a revolutionary change with the development of gene manipulation. The term gene manipulation can be applied to a variety of sophisticated in-vivo genetics as well as to in-vitro techniques. In fact, in most Western countries there is a precise *legal* definition of gene manipulation as a result of Government legislation to control it. In the United Kingdom gene manipulation is defined as 'the formation of new combinations of heritable material by the insertion of nucleic acid molecules, produced by whatever means outside the cell, into any virus, bacterial plasmid or other vector system so as to allow their incorporation into a host organism in which they do not naturally occur but in which they are capable of continued propagation.' The definitions adopted by other countries are similar and all adequately describe the subject matter of this book.

This legal definition emphasizes the propagation of foreign nucleic acid molecules (the nucleic acid is nearly always DNA) in a different, host organism. The ability to cross natural species barriers and place genes from any organism in an unrelated host organism is one important feature of gene manipulation. A second important feature is the fact that a defined and relatively small piece of DNA is propagated in the host organism. As we shall see, this has far-reaching consequences, for it is then possible to obtain a pure DNA fragment in bulk. This opens the door to a range of molecular biological opportunities including nucleotide sequence determination, site-directed mutagenesis, and manipulation of gene sequences to ensure very high level expression of an encoded polypeptide in a host organism. In addition, the DNA fragment provides a

molecular hybridization probe of absolute sequence purity, totally uncontaminated by other sequences from the donor organism.

The initial impetus for gene manipulation *in vitro* came about in the early 1970s with the simultaneous development of techniques for:

1 transformation* of *Escherichia coli*,
2 cutting and joining DNA molecules, and
3 monitoring the cutting and joining reactions.

In order to explain the significance of these developments we must first consider the essential requirements of a successful gene manipulation procedure.

The basic problems

Before the advent of modern gene manipulation methods there had been many early attempts at transforming pro- and eukaryotic cells with foreign DNA. But, in general, little progress could be made. The reasons for this are as follows. Let us assume that the exogenous DNA is taken up by the recipient cells. There are then two basic difficulties. First, where detection of uptake is dependent on gene expression failure could be due to lack of accurate transcription or translation. Second, and more importantly, the exogenous DNA may not be maintained in the transformed cells. If the exogenous DNA is integrated into the host genome, there is no problem. The exact mechanism whereby this integration occurs is not clear and it is usually a rare event. If the exogenous DNA fails to be integrated, it will probably be lost during subsequent multiplication of the host cells. The reason for this is simple. In order to be replicated, DNA molecules must contain an *origin of replication* and in bacteria and viruses there is usually only one per genome. Such molecules are called *replicons*. Fragments of DNA are not replicons and in the absence of replication will be diluted out of their host cells. It should be noted that even if a DNA molecule contains an origin of replication this may not function in a foreign host cell.

There is an additional, subsequent problem. If the early experiments were to proceed, a method was required for

* The sudden change of an animal cell possessing normal growth properties into one with many of the growth properties of the cancer cell is called *growth transformation*. Growth transformation is mentioned in Chapter 11 and should not be confused with bacterial transformation which is described here.

assessing the fate of the donor DNA. In particular, in circumstances where the foreign DNA was maintained because it had become integrated in the host DNA, a method was required for mapping the foreign DNA and the surrounding host sequences.

THE SOLUTIONS: BASIC TECHNIQUES

If fragments of DNA are not replicated, the obvious solution is to attach them to a suitable replicon. Such replicons are known as *vectors* or *cloning vehicles*. Small plasmids and bacteriophages are the most suitable vectors for they are replicons in their own right, their maintenance does not necessarily require integration into the host genome and their DNA can be isolated readily in an intact form. The different plasmids and phages which are used as vectors are described in detail in Chapters 3 and 4. Suffice it to say at this point that initially plasmids and phages suitable as vectors were only found in *Escherichia coli*.

Composite molecules in which foreign DNA has been inserted into a vector molecule are sometimes called DNA *chimaeras* because of their analogy with the Chimaera of mythology—a creature with the head of a lion, body of a goat and the tail of a serpent. The construction of such composite or *artificial recombinant* molecules has also been termed *genetic engineering* or *gene manipulation* because of the potential for creating novel genetic combinations by biochemical means. The process has also been termed *molecular cloning* or *gene cloning* because a line of genetically identical organisms, all of which contain the composite molecule, can be propagated and grown in bulk hence *amplifying* the composite molecule and *any gene product whose synthesis it directs*.

Although conceptually very simple, cloning of a fragment of foreign, or *passenger*, or *target* DNA in a vector demands that the following can be accomplished.

1 The vector DNA must be purified and cut open.

2 The passenger DNA must be inserted into the vector molecule to create the artificial recombinant. DNA joining reactions must therefore be performed. Methods for cutting and joining DNA molecules are now so sophisticated that they warrant a chapter of their own (Chapter 2).

3 The cutting and joining reactions must be readily monitored. This is achieved by the use of gel electrophoresis.

4 Finally, the artificial recombinant must be transformed into *E. coli*, or other host cell. Further details on the use of gel electrophoresis and transformation of *E. coli* are given in the

next section. As we have noted, the necessary techniques became available at about the same time and quickly led to many cloning experiments, the first of which were reported in 1972 (Jackson *et al.* 1972, Lobban & Kaiser 1973).

Agarose gel electrophoresis

The progress of the first experiments on cutting and joining of DNA molecules was monitored by velocity sedimentation in sucrose gradients. However, this has been entirely superseded

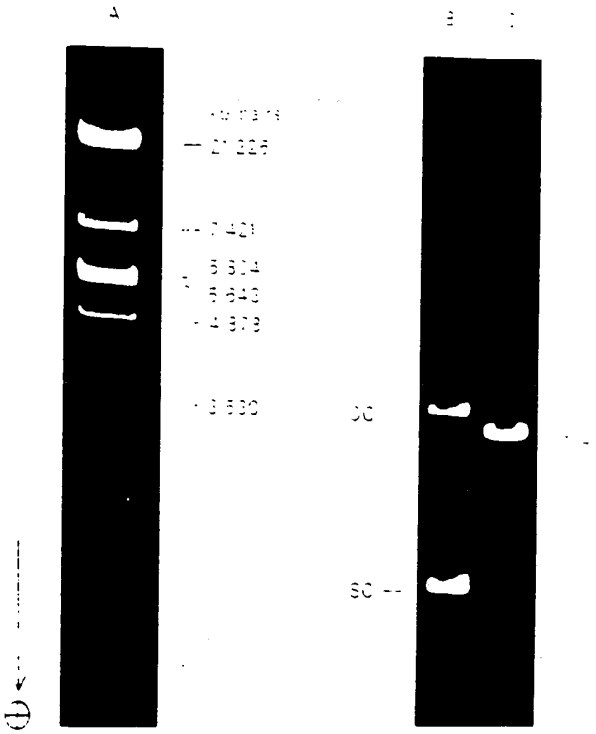


Fig. 1.1 Electrophoresis of DNA in agarose gels. The direction of migration is indicated by the arrow. DNA bands have been visualised by soaking the gel in a solution of ethidium bromide (which complexes with DNA by intercalating between stacked base pairs) and photographing the orange fluorescence which results upon ultraviolet irradiation. (A) Phage λ DNA restricted with *Eco* RI and then electrophoresed in a 1% agarose gel. The λ restriction map is given in Fig. 4.4. (B) Open circular (OC) and super-coiled (SC) forms of a plasmid of 6.4 kb pairs. Note that the compact super-coils migrate considerably faster than open circles. (C) Linear plasmid (L) DNA produced by treatment of the preparation shown in lane B with *Eco* RI for which there is a single target site. Under the conditions of electrophoresis employed here, the linear form migrates just ahead of the open-circular form.

by gel electrophoresis. Gel electrophoresis is not only used as an analytical method, it is routinely used preparatively for the purification of specific DNA fragments. The gel is composed of polyacrylamide or agarose. Agarose is convenient for separating DNA fragments ranging in size from a few hundred to about 20 000 base pairs. Polyacrylamide is preferred for smaller DNA fragments.

A gel is a complex network of polymeric molecules. DNA molecules are negatively charged and under an electric field DNA molecules migrate through the gel at rates dependent upon their sizes: a small DNA molecule can thread its way through the gel easily and hence migrates faster than a larger molecule (Fig. 1.1). Aaij and Borst (1972) showed that the migration rates of the DNA molecules were inversely proportional to the logarithms of the molecular weights. More recently, Southern (1979a,b) has shown that plotting fragment length or molecular weight against the reciprocal of mobility gives a straight line over a wider range than the semi-logarithmic plot. In any event, gel electrophoresis is frequently performed with marker DNA fragments of known size which allow accurate size determination of an unknown DNA molecule by interpolation. A particular advantage of gel electrophoresis is that the DNA bands can be readily detected at high sensitivity. The bands of DNA in the gel are stained with the intercalating dye ethidium bromide (Fig. 1.2), and as little as 0.05 μg of DNA in one band can be detected as visible fluorescence when the gel is illuminated with ultraviolet light.

In addition to resolving DNA fragments of different lengths, gel electrophoresis separates the different molecular configurations of a DNA molecule. The covalently closed circular, the nicked (relaxed) circular and linear forms of a DNA molecule have different mobilities (Fig. 1.1). Readers wishing to

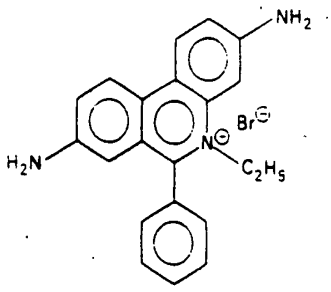


Fig. 1.2 Ethidium bromide.

know more about the factors affecting the electrophoretic mobility in agarose gels of the different conformational isomers of DNA should consult the paper by Johnson and Grossman (1977).

Southern blotting

Frequently it is necessary to know what sequences in a DNA restriction fragment are transcribed into RNA, or to be able to map sequences by hybridization to restriction fragments. Clearly it would be helpful to have a method of detecting fragments in an agarose gel that are complementary to a given RNA or DNA sequence. This can be done by slicing the gel, eluting the DNA and hybridizing to radiolabelled 'probe' DNA or RNA either in solution, or after binding the restriction fragment to filters. This method, which is time consuming and inevitably leads to some loss of resolution, has now been replaced by a neat method described by Southern (1975, 1979b). This method, often referred to as *Southern blotting* is shown in Fig. 1.3.

DNA restriction fragments in the gel are denatured by alkali treatment and the gel is then laid on top of buffer-saturated filter paper. The top surface of the gel is covered with a cellulose nitrate (often called nitrocellulose) filter and overlaid with dry filter paper. Buffer passes through the gel drawn by the dry filter paper and carries the single-stranded DNA with it. When the DNA comes into contact with the cellulose nitrate it binds to it strongly. The DNA fragments can be permanently fixed to the cellulose nitrate by baking at 80 °C. The filter can then be placed in a solution of radioactive RNA or denatured DNA which is complementary in sequence to the blot-transferred DNA. Conditions are chosen so that the radioactive nucleic acid hybridizes with complementary DNA on the cellulose nitrate. After a washing step, the regions of hybridization can be detected autoradiographically by placing the cellulose nitrate in contact with photographic film. The method is extremely sensitive and can even be used to map restriction sites around a single-copy gene sequence

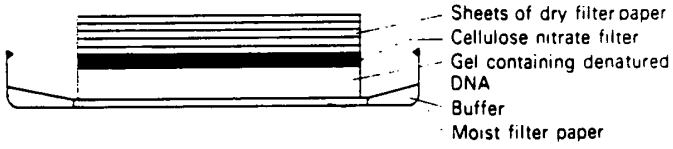


Fig. 1.3 The 'Southern blot' technique. See text for details.

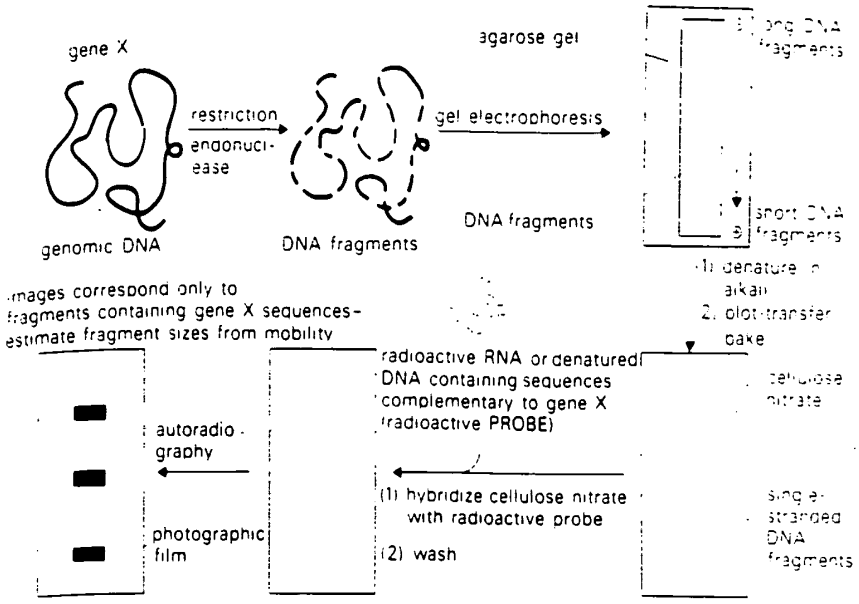


Fig. 1.4 Mapping restriction sites around a hypothetical gene sequence in total genomic DNA by the Southern blot method.

Genomic DNA is cleaved with a restriction endonuclease into hundreds of thousands of fragments of various sizes. The fragments are separated according to size by gel electrophoresis and blot-transferred on to cellulose nitrate paper. Highly radioactive RNA or denatured DNA complementary in sequence to gene X is applied to the cellulose nitrate paper bearing the blotted DNA. The radiolabelled RNA or DNA will hybridize with gene X sequences and can be detected subsequently by autoradiography, so enabling the sizes of restriction fragments containing gene X sequences to be estimated from their electrophoretic mobility. By using several restriction endonucleases singly and in combination, a map of restriction sites in and around gene X can be built up.

in a complex genome (Fig. 1.4). Since the radioactive nucleic acid is used here to search out and detect complementary sequences in the presence of a large amount of non-complementary DNA it is often referred to as the *probe*.

Northern blotting

Southern's technique has been of enormous value, but it was thought that it could not be applied directly to the blot-transfer of RNAs separated by gel electrophoresis since RNA was found not to bind to cellulose nitrate. Alwine *et al.* (1979), therefore, devised a procedure in which RNA bands are blot-transferred

from the gel onto chemically reactive paper where they are bound covalently. The reactive paper is prepared by diazotization of aminobenzyloxymethyl-paper which itself can be prepared from Whatman 540 paper by a series of uncomplicated reactions. Once covalently bound, the RNA is available for hybridization with radiolabelled DNA probes. As before, hybridizing bands are located by autoradiography. Alwine's method thus extends that of Southern and for this reason it has acquired the jargon term '*Northern*' blotting!

Because of the firm covalent binding of the RNA to the paper, such blot-transfers are reusable; the probe from previous hybridization reactions having been eluted by washing at a temperature at which hybrids are not stable. Although originally devised for the transfer of RNA bands, the chemically reactive paper is equally effective in binding denatured DNA. In fact, small DNA fragments are more efficiently transferred to the diazotized paper derivative than to nitrocellulose. Extension of the fruitful blot-transfer approach has continued and more recently it has been found that RNA bands *can* indeed be blotted directly onto nitrocellulose under appropriate conditions (Thomas 1980). Because this form of Northern blotting does not require the preparation of reactive paper, it has been widely adopted.

Boxing the compass has continued with the arrival of the term *Western* blot. This was sometimes used to describe the blotting of gel electrophoresed RNA onto nitrocellulose, as just described, but the term is now usually restricted to an application that does not involve nucleic acids, i.e. the detection of specific polypeptides by first electrophoresing them in polyacrylamide gels, blotting them onto nitrocellulose to which they bind, and then detecting them by reaction with a specific, labelled antibody.

Transformation of *E. coli*

Early attempts to achieve transformation of *E. coli* were unsuccessful and it was generally believed that *E. coli* was refractory to transformation. However, Mandel and Higa (1970) found that treatment with $CaCl_2$ allowed *E. coli* cells to take up DNA from bacteriophage λ. A few years later Cohen *et al.* (1972) showed that $CaCl_2$-treated *E. coli* cells are also effective recipients for plasmid DNA. Almost any strain of *E. coli* can be transformed with plasmid DNA, albeit with varying efficiency, whereas until recently it was thought that only *rec*BC$^-$ mutants could be transformed with linear bacterial DNA (Cosloy & Oishi

1973). More recently, Hoekstra *et al.* (1980) have shown that *rec*BC$^-$ cells can be transformed with linear DNA but the efficiency is only 10% of that in otherwise isogenic *rec*BC$^-$ cells. Transformation of *rec*BC$^-$ cells with linear DNA is only possible if the cells are rendered recombination proficient by the addition of a *sbc*A or *sbc*B mutation. The fact that the *rec*BC gene product is an exonuclease explains the difference in transformation efficiency of circular and linear DNA in *rec*BC$^-$ cells.

As will be seen from the next chapter, many bacteria contain restriction systems which can influence the efficiency of transformation. Although the complete function of these restriction systems is not known yet, one role they do play is the recognition and degradation of foreign DNA. For this reason it is usual to use a restrictionless mutant of *E. coli* as a transformable host.

Since transformation of *E. coli* is an essential step in many cloning experiments it is desirable that it be as efficient as possible. Several groups of workers have examined the factors affecting the efficiency of transformation. It has been found that *E. coli* cells and plasmid DNA interact productively in an environment of calcium ions and low temperature (0–5 °C) and that a subsequent heat shock (37–45 °C) is important, but not strictly required. Several other factors, especially the inclusion of metal ions in addition to calcium, have been shown to stimulate the process.

Hanahan (1983) has re-examined factors that affect the efficiency of transformation, and has devised a set of conditions for optimal efficiency (expressed as transformants per μg plasmid DNA) applicable to most *E. coli* K12 strains. Typically, efficiencies of 10^7 or 10^8 transformants/μg can be achieved. Large DNAs transform less efficiently, on a molar basis, than small DNAs. Even with such improved transformation procedures certain potential gene cloning experiments requiring large numbers of clones are not reliable. One approach which can be used to circumvent the problem of low transformation efficiencies is to package recombinant DNA into virus particles *in vitro*. A particular form of this approach, the use of cosmids, is described in detail in Chapter 4.

The biggest problem that prevents analysis of the transformation process is that the maximum level of transformation has not been extended beyond a few percent of the survivors of the $CaCl_2$ treatment. Weston *et al.* (1979) found that when two separate plasmids are employed simultaneously at equally saturating concentrations of DNA the total number of

transformants obtained is the same as that when one plasmid alone is used. This confirms that every transformable cell is transformed under these conditions suggesting that only a minor subpopulation of $CaCl_2$-treated cells of *E. coli* are capable of taking up and establishing plasmid molecules. This idea is supported by the work of Jones *et al.* (1981) who found that despite the fact that the transformation frequency increased 100-fold for a 10-fold increase in growth rate, the maximum number of transformable cells never exceeded 1%. Why the frequency of transformation is limited to 1% remains a mystery.

The absolute requirement for $CaCl_2$ for transformation of *E. coli* is probably due to structural alterations in the cell wall which it effects. In almost all Gram-negative bacteria which have been transformed, $CaCl_2$ treatment is a mandatory step. Only a few Gram-negative bacteria, e.g. *Haemophilus* and *Neisseria*, can be transformed without $CaCl_2$ treatment.

Transformation of other organisms

Although *E. coli* often remains the host organism of choice for cloning experiments, many other hosts are now used and with them transformation may still be a critical step. In the case of Gram-positive bacteria the two most important groups of organisms are *Bacillus* sp. and actinomycetes. That *B. subtilis* is naturally competent for transformation has been known for a long time and hence the genetics of this organism are fairly advanced. For this reason *B. subtilis* is a particularly attractive alternative prokaryotic cloning host. The significant features of transformation with this organism are detailed in Chapter 8. Of particular relevance here is that it is possible to transform protoplasts of *B. subtilis*, a technique which leads to improved transformation frequencies. A similar technique is used to transform actinomycetes and recently it has been shown that the frequency can be increased considerably by first entrapping the DNA in liposomes which then fuse with the host cell membrane. This is one aspect of transformation which undoubtedly will receive considerable attention in the near future.

In later chapters we discuss ways in which cloned DNA can be introduced into eukaryotic cells. With animal cells there is no great problem as only the membrane has to be crossed. In the case of yeast, protoplasts are required (Hinnan *et al.* 1978). With higher plants the strategy which has been considered is either to package the DNA in a plant virus or use a bacterial plant

pathogen as the donor. It has also been shown that protoplasts prepared from plant cells are competent for transformation.

OTHER TECHNIQUES

Two other basic capabilities must be included in the repertoire of the gene manipulator. The first of these is DNA sequencing. Knowledge of the sequence of a cloned DNA fragment is a prerequisite for planning any substantial manipulation of the DNA; for example, a computer search of the sequence for all known restriction endonuclease target sites will provide a complete and precise restriction map. Importantly, once the sequence has been manipulated *in vitro*, perhaps using in-vitro mutagenesis techniques, it is usually desirable to sequence the product so as to confirm that the required sequence has been obtained. The second capability is gene synthesis. Recent progress in this field has been very rapid. The synthesis of DNA molecules many hundreds of nucleotides in length is now possible. The ability to synthesize polymers up to 20 nucleotides in length quickly and easily has many important applications which are discussed in later chapters.

DNA sequencing by the Maxam and Gilbert method

This method for DNA sequencing makes use of chemical reagents to bring about base-specific cleavage of the DNA. For large-scale sequencing it is now less favoured than the enzymatic, 'dideoxy', method. However, it still finds application and illustrates principles of polyacrylamide gel electrophoresis as applied to sequence determination. Discussion of the dideoxy method is deferred until Chapter 4 because it is nearly always used in combination with phage M13 cloning vectors.

In the Maxam and Gilbert method (Maxam & Gilbert 1977), the starting point is a defined DNA restriction fragment. The DNA strand to be sequenced must be radioactively labelled at one end with a ^{32}P-phosphate group. [A detailed practical account of the entire sequencing procedure, including end-labelling methods, is available (Maxam & Gilbert 1980)]. This DNA can be either in single-stranded or duplex form. The base-specific cleavages depend upon the following points.

1 Chemical reagents have been characterized which alter one or two bases in DNA (Table 1.1). These are base-specific reactions; for example, dimethyl sulphate methylates guanine (at the N7 position).

Table 1.1 Reagents for Maxam and Gilbert DNA sequencing.

Base specificity	Base reaction	Altered base removal	Strand cleavage
1. G	dimethyl sulphate	piperidine	piperidine
2. G + A	acid	acid-catalysed depurination	piperidine
3. T + C	hydrazine	piperidine	piperidine
4. C	hydrazine + NaCl	piperidine	piperidine
5. A > C	NaOH	piperidine	piperidine

2 An altered base can then be removed from the sugar-phosphate backbone of DNA (Table 1.1).
3 The strand is cleaved with piperidine at the sugar residue lacking the base. This cleavage is dependent upon the previous step.

When each of the base-specific reagents is used in a limited reaction with end-labelled DNA, a nested set of end-labelled fragments of different lengths is generated. It is important to emphasize that the base-specific reactions are deliberately limited to give about one, or a few, cleavages per molecule. This is illustrated in Fig. 1.5 where the nested set of fragments produced by the G-specific reaction is given as an example. Sets of fragments are produced by reacting the DNA with each of the reagents separately (Table 1.1). All five reactions (1–5, Table 1.1) may be performed; the fifth reaction gives redundant

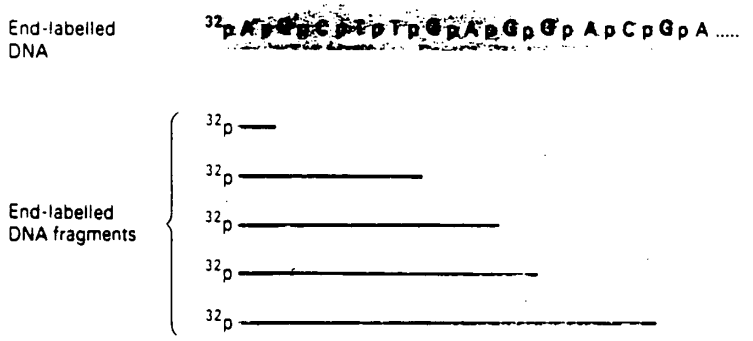


Fig. 1.5 Chemical cleavage of hypothetical DNA at G residues. A nested set of end-labelled DNA fragments is produced by limited reaction of an end-labelled DNA with G-specific reagents. Other fragments are produced but only the terminal fragments bear the label.

.Gilbert DNA sequencing.

Altered base removal	Strand cleavage
piperidine	piperidine
acid-catalysed depurination	piperidine
piperidine	piperidine
piperidine	piperidine
piperidine	piperidine

e removed from the sugar-
le 1.1).
eridine at the sugar residue
lependent upon the previous

reagents is used in a limited
a nested set of end-labelled
generated. It is important to
reactions are deliberately
cleavages per molecule. This
he nested set of fragments
is given as an example. Sets
ng the DNA with each of the
l five reactions (1–5, Table
reaction gives redundant

A p G p G p A p C p G p A ....

tical DNA at G residues. A
ients is produced by limited
h G-specific reagents. Other
minal fragments bear the label.

information but is confirmatory. These sequencing reactions are analysed by running the 4 or 5 samples side-by-side on a sequencing gel.

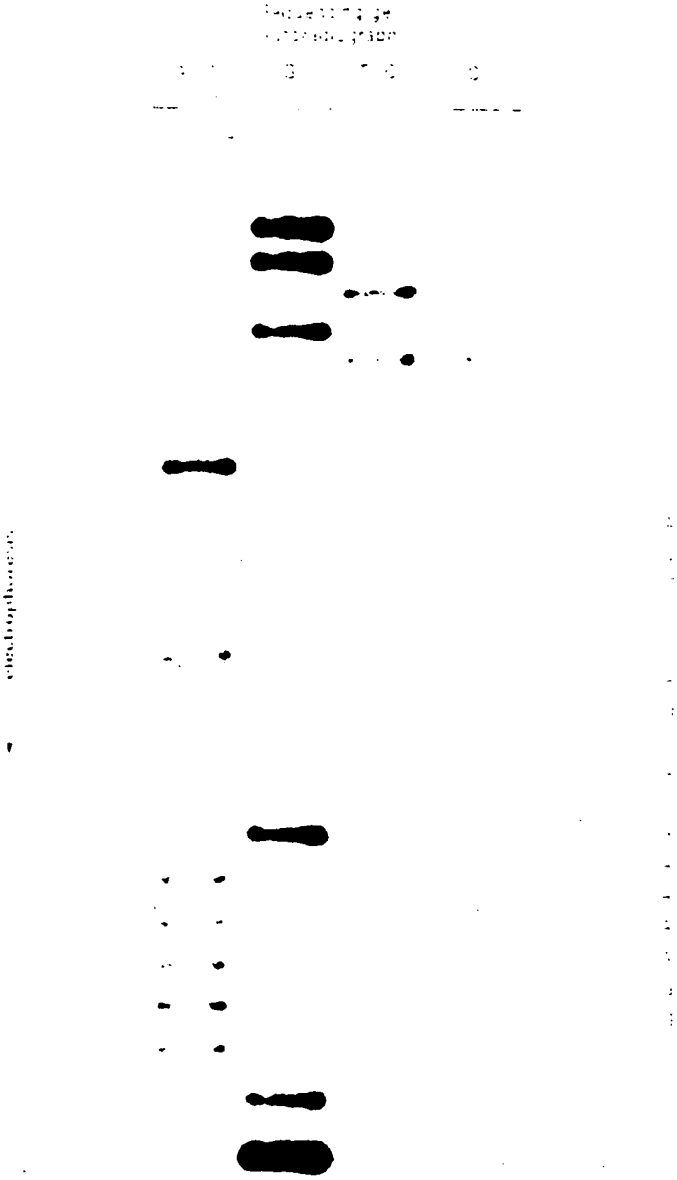


Fig. 1.6. Photograph of the Maxam and Gilbert autoradiograph (courtesy of N. Warburton).

A sequencing gel is a high resolution gel designed to fractionate single-stranded (denatured) DNA fragments on the basis of their length. They routinely contain 6% to 20% polyacrylamide and 7M urea. The urea is a denaturant whose function is to minimize DNA secondary structure effects on electrophoretic mobility. The gel is run at sufficient power to heat up to about 70 °C. This also minimizes DNA secondary structure. The labelled DNA bands obtained after such electrophoresis are revealed by autoradiography on large sheets of X-ray film. The sequence can then be read directly from the sequencing ladders in the adjacent base-specific tracks (Fig. 1.6).

The chemical synthesis of genes

The basic method of gene synthesis is the repetitive formation of an ester linkage between an activated phosphoric acid function of one nucleotide and the hydroxyl group of another nucleoside or nucleotide thus forming the characteristic phosphodiester bridge. The major problem is that deoxyribonucleotides are very reactive molecules having a primary and secondary hydroxyl group, a primary amino group and a phosphate group. Consequently, blocking and deblocking procedures are required and the chemistry involved must not result in scission or alteration of the phosphodiester backbone, the furanose rings, the sugar-purine/pyrimidine bond or the bases themselves—a tall order indeed! The various methods used to synthesize oligonucleotides are given in outline form below. The detailed chemistry is beyond the scope of this book. The reader wishing more information should consult the reviews of Caruthers *et al.* (1982) and Narang (1983).

The original method for synthesizing oligonucleotides is the so-called phosphodiester approach. This method is shown in Fig. 1.7. The 5'-phosphate of one nucleotide, after suitable protection of other functional groups, is condensed with the 3'-hydroxyl of another protected nucleoside or nucleotide using dicyclohexylcarbodiimide or an arylsulphonyl chloride. The chain can then be elongated by further step-wise condensations. The problems with the phosphodiester approach are long reaction times, rapidly decreasing yields as the chain length increases and time-consuming purification procedures. Nevertheless, the method was used successfully to synthesize a biologically active tRNA gene (Khorana 1979).

The phosphotriester approach solves some of the problems of the phosphodiester method by blocking each internucleotide

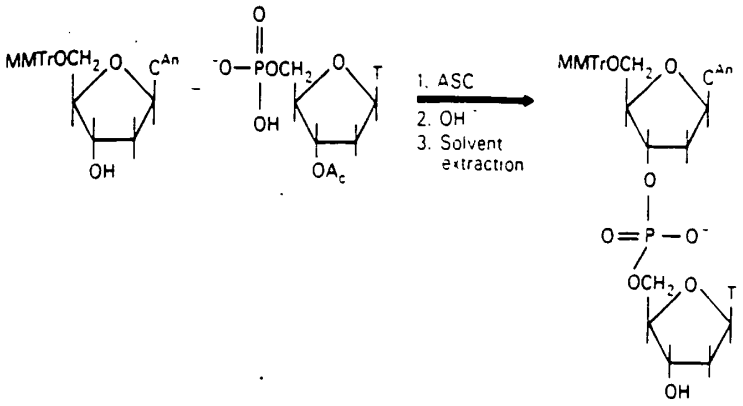


Fig. 1.7 The phosphodiester method of synthesizing oligonucleotides. The various blocking groups shown are: MMTr, monomethoxytrityl; Ac, acetyl; An, anisoyl.

phosphodiester function during the course of building a defined sequence. The basic principle of the method is the use of a totally protected mononucleotide containing a fully masked 3'-phosphotriester group (Fig. 1.8). Using this methodology Edge *et al.* (1981) synthesized 67 oligonucleotides of chain length 10–20 and spliced them together to generate a 517bp interferon-alpha gene.

More recently the phosphite-triester method has been introduced and this makes use of the extreme reactivity of phosphite reagents. In this way the two building blocks are joined in a few minutes compared with the hours required by the phosphotriester approach. The basic principles of the method are shown in Fig. 1.9. Although the phosphite-triester method will work in solution for the condensation of 3 or 4 nucleotides, construction of large oligonucleotides requires the 3' end of the desired oligonucleotide to be coupled to an insoluble support. Fortunately, immobilization simplifies manipulative procedures and cuts out time-consuming purification steps following each cycle of condensation. Appropriately blocked mononucleotides are added sequentially and reagents, starting materials and by-products removed by filtration. At the conclusion of the synthesis the deoxyoligonucleotide is chemically freed of blocking groups, hydrolysed from the support and purified by electrophoresis or HPLC. By

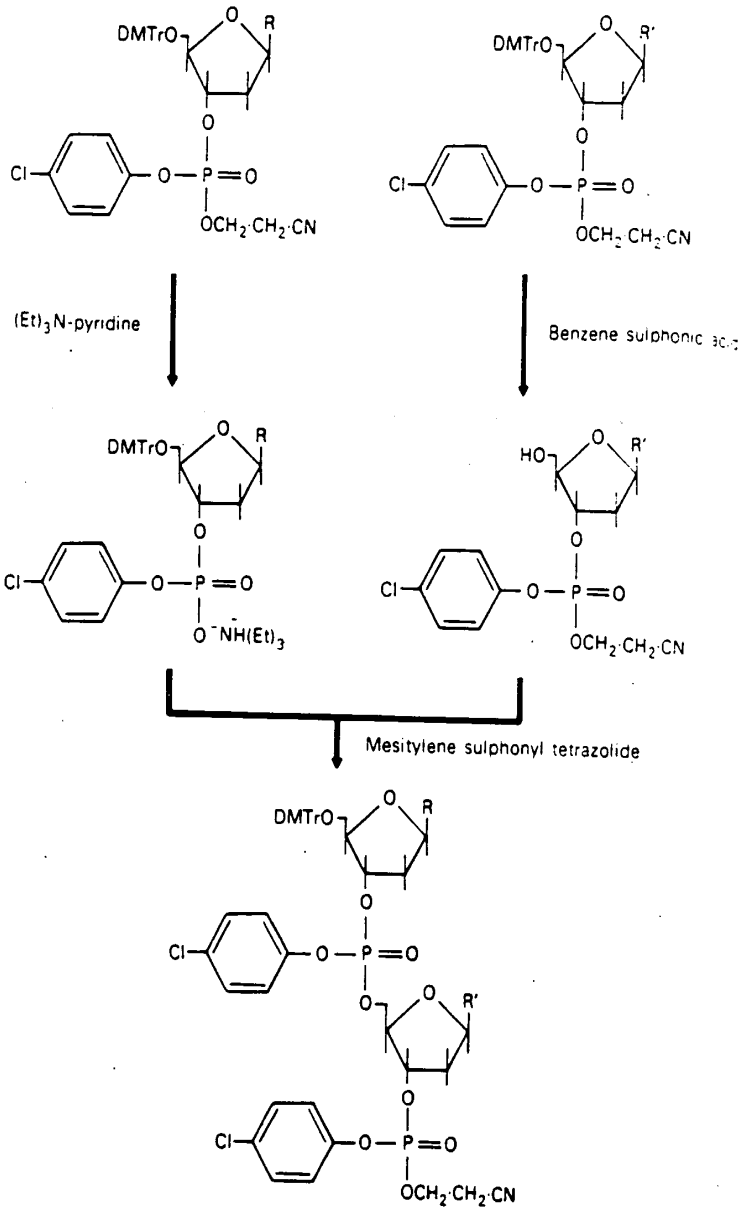
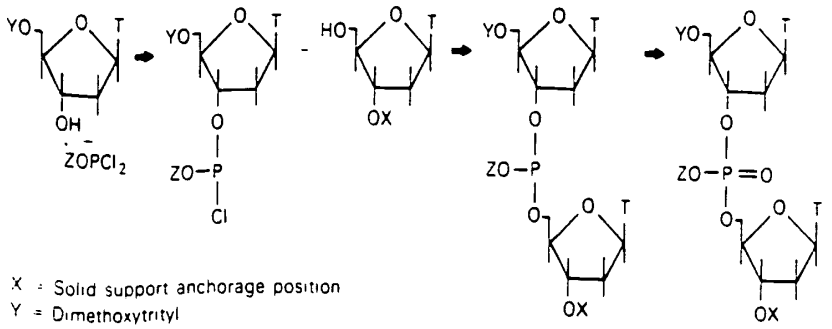


Fig. 1.8 The phosphotriester method of synthesizing oligonucleotides. The letter R or R' represent any one of the four bases: adenine, guanine, cytosine or thymine. DMTr represents the dimethoxytrityl blocking group.



X = Solid support anchorage position
Y = Dimethoxytrityl
Z = Methyl

Fig. 1.9 The phosphite-triester method of synthesizing oligonucleotides.

immobilizing the polymer support carrying the initiating deoxynucleotide in a column, the filtration steps can be replaced by a simple washing procedure and this lends itself to a fully automatic synthesis. Automatic gene synthesizers consist simply of reagent reservoirs whose contents are added to or removed from the immobilized protected oligonucleotide via valves controlled by a microcomputer.

Chapter 2 Cutting and Joining DNA Molecules

CUTTING DNA MOLECULES

It is worth recalling that prior to 1970 there was simply no method available for cutting a duplex DNA molecule into discrete fragments. DNA biochemistry was circumscribed by this impasse. It became apparent that the related phenomena of host-controlled restriction and modification might lead towards a solution to the problem when it was discovered that restriction involves specific endonucleases. The favourite organism of molecular biologists, *E. coli* K12, was the first to be studied in this regard, but turned out to be an unfortunate choice. Its endonuclease is perverse in the complexity of its behaviour. The breakthrough in 1970 came with the discovery in *Haemophilus influenzae* of an enzyme that behaves more simply. Present-day DNA technology is totally dependent upon our ability to cut DNA molecules at specific sites with restriction endonucleases. An account of host-controlled restriction and modification therefore forms the first part of this chapter.

Host-controlled restriction and modification

Host-controlled restriction and modification are most readily observed when bacteriophages are transferred from one bacterial host strain to another. If a stock preparation of phage λ, for example, is made by growth upon *E. coli* strain C and this stock is then titred upon *E. coli* C and *E. coli* K, the titres observed on these two strains will differ by several orders of magnitude, the titre on *E. coli* K being the lower. The phage are said to be *restricted* by the second host strain (*E. coli* K). When those phage that do result from the infection of *E. coli* K are now replated on *E. coli* K they are no longer restricted; but if they are first cycled through *E. coli* C they are once again restricted when plated upon *E. coli* K (Fig. 2.1). Thus the efficiency with which phage λ plates upon a particular host strain depends upon the strain on which it was last propagated. This non-heritable

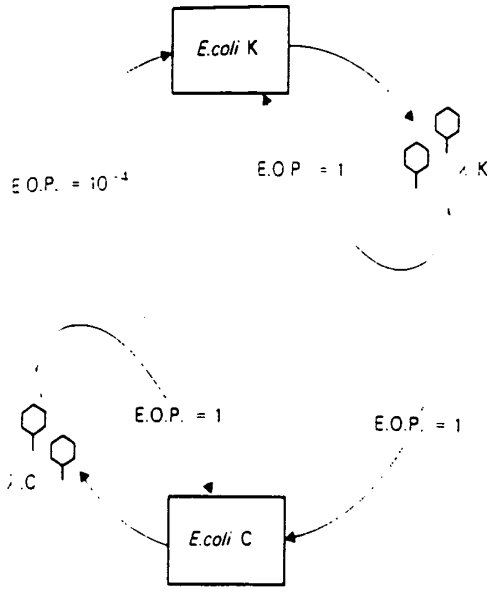


Fig. 2.1 Host-controlled restriction and modification of phage λ in *E. coli* strain K, analysed by efficiency of plating (E.O.P.). Phage propagated by growth on strains K or C (i.e. λ.K or λ.C) have E.O.P.s on the two strains as indicated by arrows. *E. coli* C has no known restriction and modification system.

change conferred upon the phage by the second host strain (*E. coli* K) that allows it to be replated on that strain without further restriction is called modification.

The restricted phages adsorb to restrictive hosts and inject their DNA normally. When the phage are labelled with ^{32}P it is apparent that their DNA is degraded soon after injection (Dussoix & Arber 1962) and the endonuclease that is primarily responsible for this degradation is called a *restriction endonuclease* or restriction enzyme (Lederberg & Meselson 1964). The restrictive host must, of course, protect its own DNA from the potentially lethal effects of the restriction endonuclease and so its DNA must be appropriately modified. Modification involves methylation of certain bases at a very limited number of sequences within DNA which constitute the recognition sequences for the restriction endonuclease. This explains why phage that survive one cycle of growth upon the restrictive host can subsequently reinfect that host efficiently; their DNA has been replicated in the presence of the modifying methylase and so it, like the host DNA, becomes methylated and protected from the restriction system.

Although phage infection has been chosen as our example to illustrate restriction and modification, these processes can occur whenever DNA is transferred from one bacterial strain to another. Conjugation, transduction, transformation and transfection are all subject to the constraint of host-controlled restriction. The genes that specify host-controlled restriction and modification systems may reside upon the bacterial chromosome itself or may be located on a plasmid or prophage such as P1.

The restriction endonuclease of *E. coli* K was the first to be isolated and studied in detail. Meselson and Yuan (1968) achieved this by devising an ingenious assay in which a fractionated cell extract was incubated with a mixture of unmodified and modified phage λ DNAs which were differentially radiolabelled—one with ^{3}H, the other with ^{32}P—so that they could be distinguished. After incubation, the DNA mixture was analysed by sedimentation through a sucrose gradient where the appearance of degraded unmodified DNA in the presence of undegraded modified DNA indicated the activity of restriction endonuclease.

The enzyme from *E. coli* K, and the similar one from *E. coli* B, were found to have unusual properties. In addition to magnesium ions, they require the cofactors ATP and S-adenosyl-methionine, and DNA degradation *in vitro* is accompanied by hydrolysis of the ATP in amounts greatly exceeding the stoichiometry of DNA breakage (Bickle *et al.* 1978). In addition, the enzymes are now known to interact with an unmodified *recognition* sequence in duplex DNA and then surprisingly, to track along the DNA molecule. After travelling for a distance corresponding to between 1000 and 5000 nucleotides the enzyme cleaves one strand only of the DNA at an apparently random site, and makes a gap about 75 nucleotides in length by releasing acid-soluble oligonucleotides. There is no evidence that the enzyme is truly catalytic, and having acted once in this way a second enzyme molecule is required to complete the double-strand break (Rosamond *et al.* 1979). Enzymes with these properties are now known as type I restriction endonucleases. Like all restriction endonucleases they recognize specific nucleotide sequences. However, they are not particularly useful for gene manipulation since their cleavage sites are non-specific. Their biochemistry still presents many puzzles; for instance, the precise role of S-adenosyl-methionine remains unclear.

While these bizarre properties of type I restriction enzymes were being unravelled, a restriction endonuclease from *H. influenzae* Rd was discovered (Kelly & Smith 1970, Smith

& Wilcox 1970) that was to become the prototype of a large number of restriction endonucleases—now known as type II enzymes—that have none of the unusual properties displayed by type I enzymes and which are fundamentally important in the manipulation of DNA. The type II enzymes recognize a particular target sequence in a duplex DNA molecule and break the polynucleotide chains within, or near to, that sequence to give rise to discrete DNA fragments of defined length and sequence. In fact, the activity of these enzymes is often assayed and studied by gel electrophoresis of the DNA fragments which they generate (see Fig. 1.1). As expected, digests of small plasmid or viral DNAs give characteristic simple DNA band patterns.

Very many type II restriction endonucleases have now been isolated from a wide variety of bacteria. In a recent review, Roberts (1984) lists 475 enzymes that have been at least partially characterized, and the number continues to grow as more bacterial genera are surveyed for their presence. It is worth noting that many so-called restriction endonucleases have not formally been shown to correspond with any genetically identified restriction and modification system of the bacteria from which they have been prepared: it is usually assumed that a site-specific endonuclease which is inactive upon host DNA and active upon exogenous DNA is, in fact, a restriction endonuclease.

Nomenclature

The discovery of a large number of restriction enzymes called for a uniform nomenclature. A system based upon the proposals of Smith and Nathans (1973) has been followed for the most part. The proposals were as follows.

1 The species name of the host organism is identified by the first letter of the genus name and the first two letters of the specific epithet to form a three-letter abbreviation in italics: for example, *Escherichia coli* = *Eco* and *Haemophilus influenzae* = *Hin*.

2 Strain or type identification is written as a subscript, e.g. Eco_K. In cases where the restriction and modification system is genetically specified by a virus or plasmid, the abbreviated species name of the host is given and the extrachromosomal element is identified by a subscript, e.g. Eco_{P1}, Eco_{R1}.

3 When a particular host strain has several different restriction and modification systems, these are identified by Roman numerals, thus the systems from *H. influenzae* strain Rd would

Table 2.1 Target sites for some restriction endonucleases. A comprehensive list is given in Appendix 2.

Anabaena variabilis	*Ava* I	C↓(C/T)CG(A/G)G	
Bacillus amyloliquefaciens H	*Bam* HI	G↓GATCC	
Bacillus globigii	*Bgl* II	A↓GATCT	
Escherichia coli RY13	*Eco* RI	G↓AA*TTC	1,4
Escherichia coli R245	*Eco* RII	↓CC(A/T)GG	2
Haemophilus aegyptius	*Hae* III	GG↓C*C	
Haemophilus gallinarum	*Hga* I	GACGC	3
Haemophilus haemolyticus	*Hha* I	GC*G↓C	
Haemophilus influenzae Rd	*Hind* II	GT(C/T)↓(A/G)*AC	
	Hind III	A*↓AGCTT	
Haemophilus parainfluenzae	*Hpa* I	GTT↓AAC	
	Hpa II	C↓C*GG	
Klebsiella pneumoniae	*Kpn* I	GGTAC↓C	
Moraxella bovis	*Mbo* I	↓GATC	
Providencia stuartii	*Pst* I	CTGCA↓G	
Serratia marcescens	*Sma* I	CCC↓GGG	
Streptomyces stanford	*Sst* I	GAGCT↓C	
Xanthomonas malvacearum	*Xma* I	C↓CCGGG	

Source: Roberts (1984). Recognition sequences are written from 5′ → 3′, only one strand being given, and the point of cleavage is indicated by an arrow. Bases written in parentheses signify that either base may occupy that position. Where known, the base modified by the corresponding specific methylase is indicated by an asterisk.
A* is N⁶-methyladenine, C* is 5-methylcytosine.

Notes
1, 2. The names of these two enzymes are anomalous. The genes specifying the enzymes are borne on two Resistance Transfer Factors which have been classified separately. Hence RI and RII.
3. *Hga* I is a Type II restriction endonuclease, cleaving as indicated:

5′ GACGCNNNNN↓
3′ CTGCGNNNNN NNNNN↑

where N is any nucleotide.
4. Under certain conditions (low ionic strength, alkaline pH or 50% glycerol) the *Eco* RI specificity is reduced so that only the internal tetranucleotide sequence of the canonical hexanucleotide is necessary for recognition and cleavage. This is so-called *Eco* RI* (RI-star) activity. It is inhibited by parachloromercuribenzoate, whereas *Eco* RI activity is insensitive (Tikchonenko *et al.* 1978). Many other enzymes exhibit star activity, i.e. reduced specificity, under suboptimal conditions.

be *Hin*$_d$I, *Hin*$_d$II, *Hin*$_d$III, etc. These Roman numerals should not be confused with those in the classification of restriction enzymes into type I, etc.

All restriction enzymes have the general name endonuclease R, but, in addition, carry the system name, e.g. endonuclease R.*Hin*$_d$III. Similarly, modification enzymes are named methylase M followed by the system name. The modification enzyme from *H. influenzae* Rd corresponding to endonuclease R.*Hin*$_d$III is designated methylase M.*Hin*$_d$III.

In practice this system of nomenclature has been simplified further:

(a) subscripts are typographically inconvenient: the whole abbreviation is now usually written on the line;

(b) where the context makes it clear that restriction enzymes only are involved, the designation endonuclease R. is omitted. This is the system used in Table 2.1, which lists some of the more commonly used restriction endonucleases. A more extensive list is given in Appendix 2.

Target sites

The vast majority of, but not all, type II restriction endonucleases recognize and break DNA within particular sequences of tetra-, penta-, hexa- or hepta-nucleotides which have an axis of *rotational symmetry*; for example, *Eco* RI cuts at the positions indicated by arrows in the sequence

axis of symmetry

5′—G A Ă T T C—
3′—C T T A A G—

giving rise to termini bearing 5′-phosphate and 3′-hydroxyl groups. Such sequences are sometimes said to be *palindromic* by analogy with words that read alike backwards and forwards. (However, this term has also been applied to sequences such as

5′—A G C C G A—
3′—T C G G C T—

which are palindromic *within one strand*, yet do not have an axis of rotational symmetry.) If the sequence is modified by methylation so that 6-methyladenine residues are found at *one* or *both* of the positions indicated by asterisks then the sequence is resistant to endonuclease R.*Eco* RI. The resistance of the half-methylated site protects the bacterial host's own duplex DNA from attack immediately after semi-conservative repli-

cation of the fully methylated site until the modification methylase can once again restore the daughter duplexes to the fully methylated state.

We can see that *Eco* RI makes single-strand breaks four nucleotide pairs apart in the opposite strands of its target sequence, and so generates fragments with protruding 5′-termini. These DNA fragments can associate by hydrogen bonding between overlapping 5′-termini, or the fragments can circularize by intramolecular reaction, and for this reason the fragments are said to have *sticky* or *cohesive* ends (Fig. 2.2). In principle, DNA fragments from diverse sources can be joined by means of the cohesive ends, and it is possible, as we shall see later, to seal the remaining nicks in the two strands to form an intact *artificially recombinant* duplex DNA molecule.

It is clear from Table 2.1 that not all type II enzymes cleave their target sites like *Eco* RI. Some enzymes (e.g. *Pst* I) produce fragments bearing 3′-cohesive ends. Others (e.g. *Hae* III) make even cuts giving rise to flush- or blunt-ended fragments with no cohesive end at all. Some enzymes recognize tetranucleotide sequences, others recognize longer sequences, and this of course determines the average fragment length produced. We would expect any particular tetranucleotide target to occur about once every 4^4 (i.e. 256) nucleotide pairs in a long random

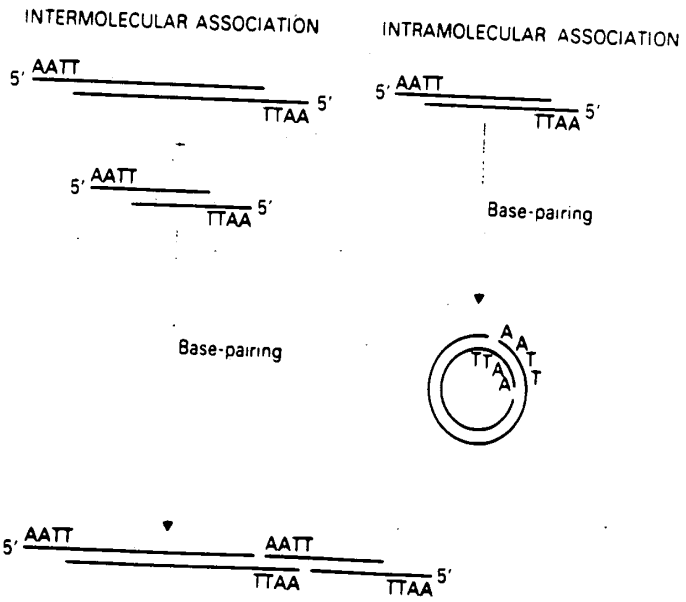


Fig. 2.2 Cohesive ends of DNA fragments produced by digestion with *Eco* RI.

DNA sequence, assuming all bases are equally frequent. Any particular hexanucleotide target would be expected to occur once in every 4^6 (i.e. 4096) nucleotide pairs. Some enzymes (e.g. *Sau* 3AI) recognize a tetranucleotide sequence that is included within the hexanucleotide sequence recognized by a different enzyme (e.g. *Bam* HI). The cohesive termini produced by these enzymes are such that fragments produced by *Sau* 3AI will cohere with those produced by *Bam* HI. If the fragments are then covalently joined, the 'hybrid site' so produced will be once again sensitive to *Sau* 3AI, but may not constitute a target for *Bam* HI; this will depend upon the nucleotides adjacent to the original *Sau* 3AI site (Fig. 2.3). Several other combinations of enzymes have this property.

From Table 2.1 we can also see that *Hind* II, the first type II enzyme to be discovered, is an example of an enzyme recognizing a sequence with some ambiguity; in this case all three sequences corresponding to the structure given in Table 2.1 are substrates. There are also several known examples of enzymes from different sources which recognize the same target. They are *isoschizomers*. Some pairs of isoschizomers cut their target at different places (e.g. *Sma* I, *Xma* I).

In our discussion of the phenomena of restriction and modification of phage λ by *E. coli* K, we saw that methylation was the basis of modification in that system. In the wide variety of type II restriction enzymes now known there are some curious and useful examples of the influence of methyl groups at restriction sites. The enzymes *Hpa* II and *Msp* I are isoschizomers with the target sequence CCGG. *Hpa* II will not cut the target when it contains 5-methylcytosine as indicated by the asterisk CC*GG (indeed this is the product of M. *Hpa* II). However, *Msp* I is known to be indifferent to methylation at

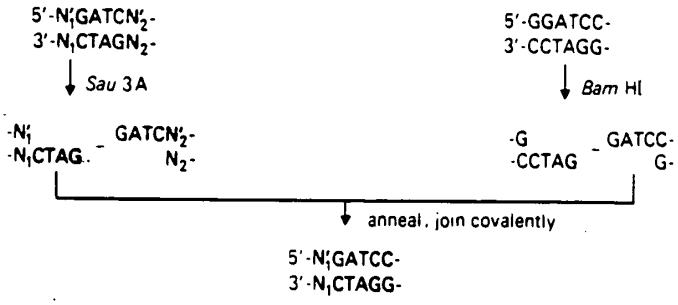


Fig. 2.3 Production of a hybrid site by cohesion of complementary sticky ends generated by *Sau* 3A and *Bam* HI.

this nucleotide: it cleaves whether or not this C residue is methylated. Now it has been found that over 90% of the methyl groups in genomic DNA of many animals, including vertebrates and echinoderms, occur as 5-methylcytosine in the sequence CG. Many of these methyl groups occur at *Msp* I sites, and their presence can be detected by comparing digests of the DNA generated by *Hpa* II and *Msp* I. Indeed methylation at *Msp* I sites around a single gene can be investigated in detail by combining the use of these two enzymes with the Southern-blot technique (Bird & Southern 1978, Razin & Riggs 1980).

Recently a new kind of type II restriction endonuclease has been identified. Enzymes of this type make breaks in the two strands at *measured distances* to one side of their asymmetric target sequence (e.g. *Hga* I; see Table 2.1). Finally, type III restriction enzymes have been classified as those which cleave DNA at well-defined sites, require ATP and Mg^{2+}, but which have only a partial requirement for (i.e. are stimulated by) S-adenosylmethionine. In these respects they have properties intermediate between type I and type II enzymes (Table 2.2).

The wide variety of properties exhibited by restriction endonucleases described in the preceding paragraphs has provided great scope for ingenious and resourceful gene manipulators. This will be apparent from examples in following chapters.

What is the function of restriction endonucleases *in vivo*? Clearly host-controlled restriction acts as a mechanism by which bacteria distinguish self from non-self. It is analogous to an immunity system. Restriction is moderately effective in preventing infection by some bacteriophages. It may be for this reason that the T-even phages (T2, T4 and T6) have evolved with glucosylated hydroxymethylcytosine residues replacing cytosine in their DNA, so rendering it resistant to many restriction endonucleases. The restriction and glucosylation modification of T-even phage DNA is beyond the scope of this book. For a detailed discussion the reader is referred to Kornberg (1980). However, it is worth noting that a mutant strain of T4 is available which does have cytosine residues in its DNA and is therefore amenable to conventional restriction methodology (Velten *et al.* 1976, Murray *et al.* 1979, Krisch & Selzer 1981). As an alternative to the unusual DNA structure of the T-even phages, other mechanisms appear to have evolved in T3 and T7 for overcoming restriction *in vivo* (Spoerel *et al.* 1979). In spite of this evidence we may be mistaken in concluding that immunity to phage infection is the sole or main function of restriction endonucleases in nature.

Table 2.2 Characteristics of restriction endonucleases (Yuan 1981, Ida *et al.* 1982, Hadi *et al.* 1982).

Table 2.2 Characteristics of restriction endonucleases (Yuan 1981, Iida *et al.* 1982, Hadi *et al.* 1982).

	Type I	Type II	Type III
1. Restriction and modification activities	Single multifunctional enzyme	Separate endonuclease and methylase	Separate enzymes with a subunit in common
2. Protein structure of restriction endonuclease	3 different subunits	Simple	2 different subunits
3. Requirements for restriction	ATP, Mg^{2+} S-adenosyl-methionine	Mg^{2+}	ATP, Mg^{2+} (S-adenosyl-methionine)
4. Sequence of host specificity sites	*Eco* B: $TGAN_8TGCT$ *Eco* K: $AACN_6GTGC$	rotational symmetry	*Eco* P1: ACACC *Eco* P15: CAGCAG
5. Cleavage sites	Possibly random, at least 1000 bp from host specificity site	At or near host specificity site	24–26 bp to 3′ of host specificity site
6. Enzymatic turnover	No	Yes	Yes
7. DNA translocation	Yes	No	No
8. Site of methylation	Host specificity site	Host specificity site	Host specificity site

N = any nucleotide

Mechanical shearing of DNA

In addition to digesting DNA with restriction endonucleases to produce discrete fragments, there are a variety of treatments which result in non-specific breakage. Non-specific endonucleases and chemical degradation can be used but the only method that has been much applied to gene manipulation involves mechanical shearing.

The long, thin threads which constitute duplex DNA molecules are sufficiently rigid to be very easily broken by shear forces in solution. Intense sonication with ultrasound can reduce the length to about 300 nucleotide pairs. More controlled shearing can be achieved by high-speed stirring in a blender. Typically, high mol. wt DNA is sheared to a population of molecules with a mean size of about 8 kb pairs by stirring at 1500 rev/min for 30 min (Wensink *et al.* 1974). Breakage occurs essentially at random with respect to DNA sequence. The termini consist of short single-stranded regions which may have to be taken into account in subsequent joining procedures.

JOINING DNA MOLECULES

Having described the methods available for cutting DNA molecules we must consider the ways in which DNA fragments can be joined to create artificially recombinant molecules. There are currently three methods for joining DNA fragments *in vitro*. The first of these capitalizes on the ability of DNA ligase to join covalently the annealed cohesive ends produced by certain restriction enzymes. The second depends upon the ability of DNA ligase from phage T4-infected *E. coli* to catalyse the formation of phosphodiester bonds between blunt-ended fragments. The third utilizes the enzyme terminal deoxynucleotidyltransferase to synthesize homopolymeric 3'-single-stranded tails at the ends of fragments. We can now look at these three methods a little more deeply.

DNA ligase

E. coli and phage T4 encode an enzyme, DNA ligase, which seals single-stranded nicks between adjacent nucleotides in a duplex DNA chain (Olivera *et al.* 1968, Gumport & Lehman 1971). Although the reactions catalysed by the enzymes of *E. coli* and T4-infected *E. coli* are very similar, they differ in their cofactor requirements. The T4 enzyme requires ATP, whilst the *E. coli* enzyme requires NAD^+. In each case the cofactor is split

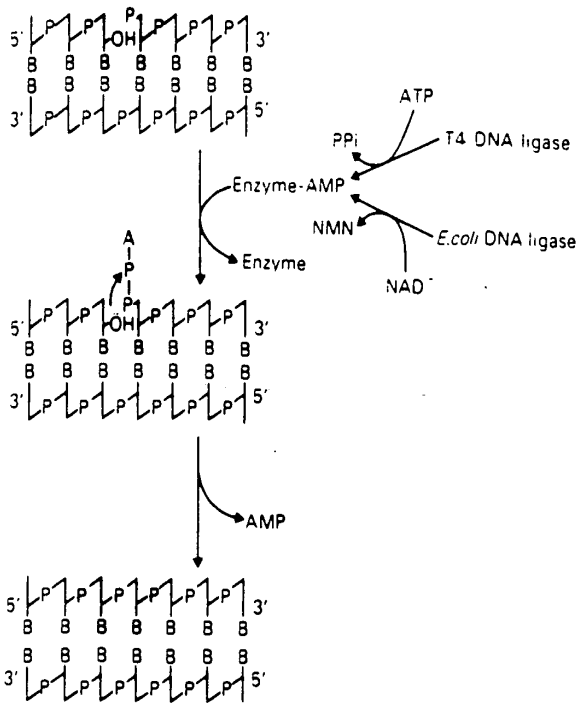


Fig. 2.4. Action of DNA ligase. An enzyme-AMP complex binds to a nick bearing 3'-OH and 5'-P groups. The AMP reacts with the phosphate group. Attack by the 3'-OH group on this moiety generates a new phosphodiester bond which seals the nick.

and forms an enzyme-AMP complex. The complex binds to the nick, which must expose a 5'-phosphate and 3'-OH group, and makes a covalent bond in the phosphodiester chain as shown in Fig. 2.4.

When termini created by a restriction endonuclease which creates cohesive ends associate, the joint has nicks a few base pairs apart in opposite strands. DNA ligase can then repair these nicks to form an intact duplex. This reaction, performed *in vitro* with purified DNA ligase, is fundamental to many gene manipulation procedures, such as that shown in Fig. 2.5.

The optimum temperature for ligation of nicked DNA is 37°C, but at this temperature the hydrogen-bonded joint between the sticky ends is unstable. *Eco* RI-generated termini associate through only four A.T base pairs and these are not sufficient to resist thermal disruption at such a high temperature. The optimum temperature for ligating the cohesive termini is therefore a compromise between the rate of enzyme

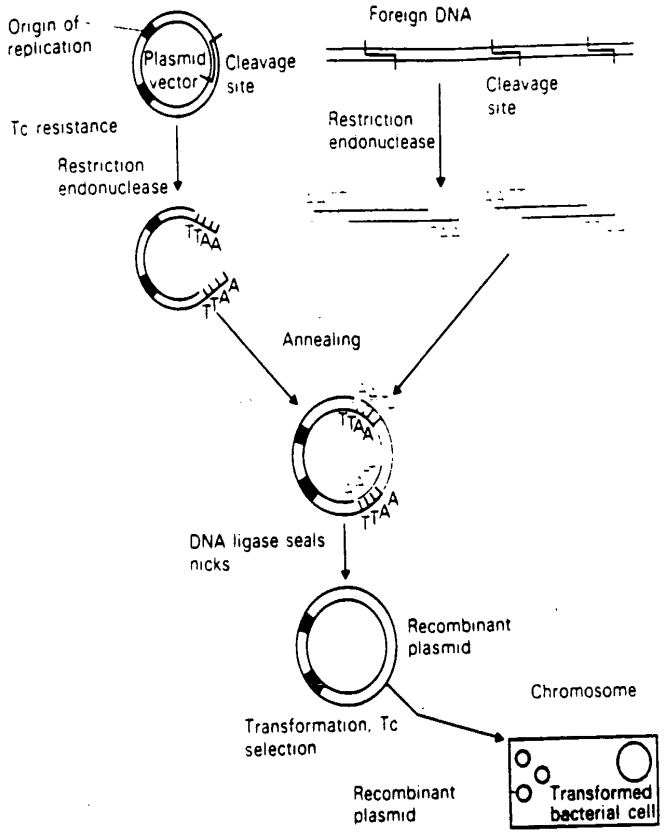


Fig. 2.5 Use of DNA ligase to create a covalent DNA recombinant joined through association of termini generated by *Eco* RI.

action and association of the termini, and has been found by experiment to be in the range 4–15 °C (Dugaicyzk *et al.* 1975, Ferretti & Sgaramella 1981).

The ligation reaction can be performed so as to favour the formation of recombinants. First, the population of recombinants can be increased by performing the reaction at a high DNA concentration; in dilute solutions *circularization* of linear fragments is relatively favoured because of the reduced frequency of intermolecular reactions. Second, by treating linearized plasmid vector DNA with alkaline phosphatase to remove 5'-terminal phosphate groups, both recircularization and plasmid dimer formation are prevented (Fig. 2.6). In this case, circularization of the vector can occur only by insertion of non-phosphatase-treated foreign DNA which provides one 5'-

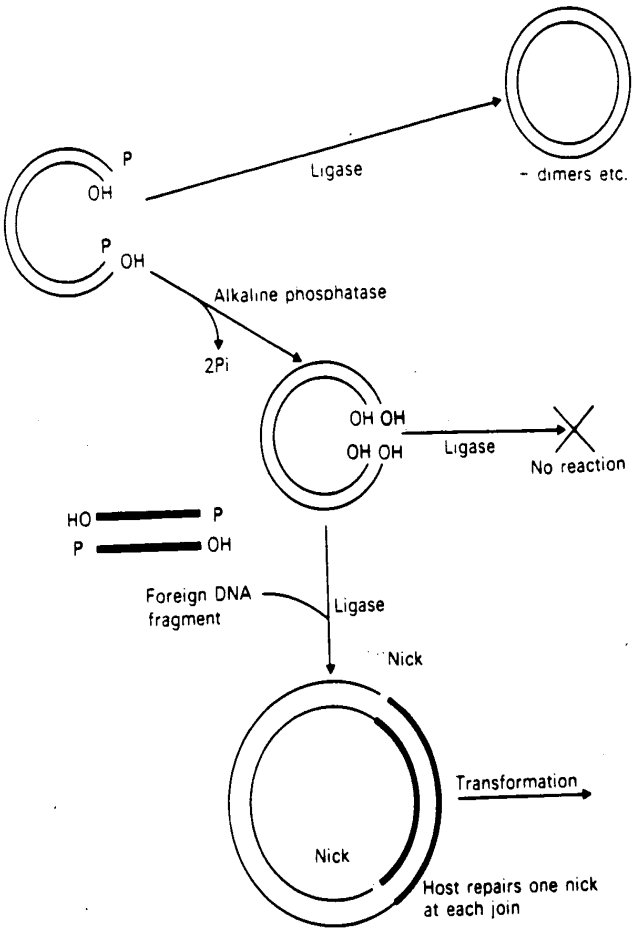


Fig. 2.6 Application of alkaline phosphatase treatment to prevent recircularization of vector plasmid without insertion of foreign DNA.

terminal phosphate at each join. One nick at each join remains unligated, but after transformation of host bacteria, cellular repair mechanisms reconstitute the intact duplex.

Joining DNA fragments with cohesive ends by DNA ligase is a relatively efficient process which has been used extensively to create artificial recombinants. A modification of this procedure depends upon the ability of T4 DNA ligase to join blunt-ended DNA molecules (Sgaramella 1972). The *E. coli* DNA ligase will not catalyse blunt ligation except under special reaction conditions of macromolecular crowding (Zimmerman *et al.* 1983). Blunt ligation is most usefully applied to joining blunt-ended fragments via *linker* molecules; for example,

Scheller *et al.* (1977) have synthesized self-complementary decameric oligonucleotides which contain sites for one or more restriction endonucleases. One such molecule is shown in Fig. 2.7. The molecule can be ligated to both ends of the foreign DNA to be cloned, and then treated with restriction endonuclease to produce a sticky-ended fragment which can be incorporated into a vector molecule that has been cut with the same restriction endonuclease. Insertion by means of the linker creates restriction enzyme target sites at each end of the foreign DNA and so enables the foreign DNA to be excised and recovered after cloning and amplification in the host bacterium.

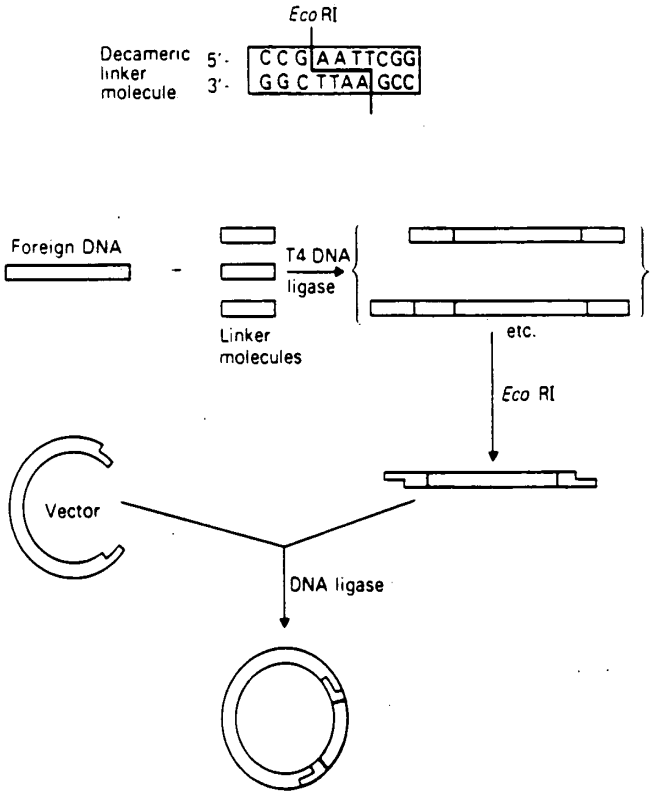


Fig. 2.7 A decameric linker molecule containing an *Eco* RI target site is joined by T4 DNA ligase to both ends of flush-ended foreign DNA. Cohesive ends are then generated by *Eco* RI. This DNA can then be incorporated into a vector that has been treated with the same restriction endonuclease.

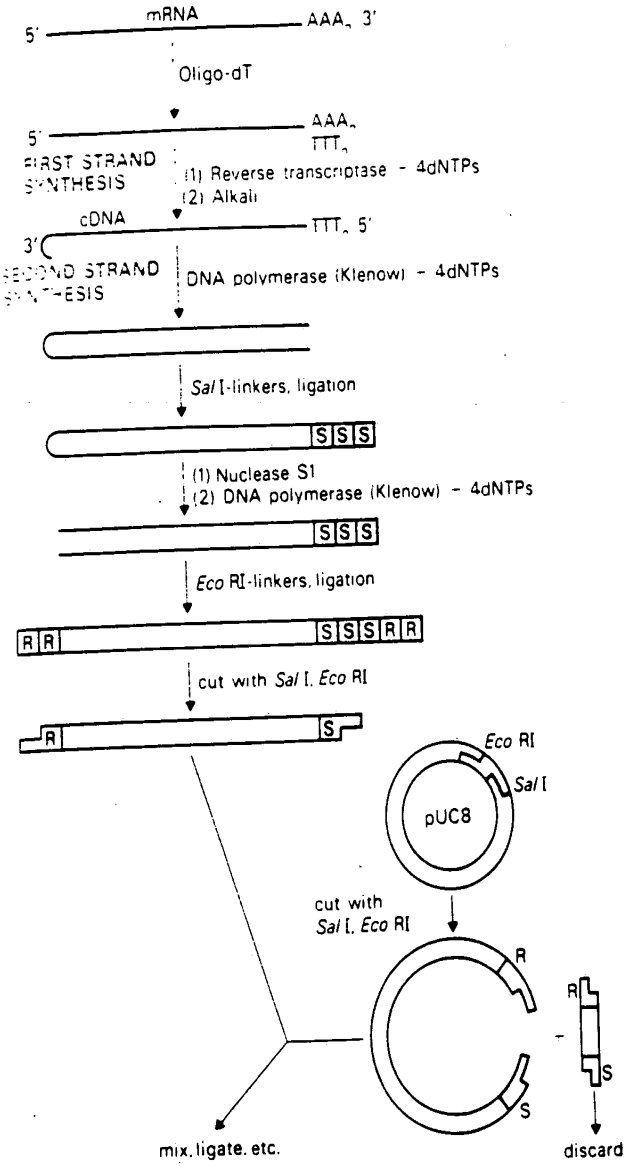


Fig. 2.8 Double-linkers. mRNA is copied into double-stranded cDNA and *Sal* I-linkers (S) are added. The hairpin loop formed by self-priming of second strand synthesis is removed by nuclease S1. In this procedure, any raggedness left by nuclease S1 (i.e. short single-strand projections at the terminus) is removed by polishing with Klenow polymerase. *Eco* RI-linkers are then ligated to the duplex molecule, and cohesive termini revealed by restriction with *Sal* I and *Eco* RI. The cDNA plus linkers is then ligated into a vector cut with the same two enzymes.

Double-linkers

Plasmid vectors have been derived which contain a set of closely clustered cloning sites. An example of such a vector is pUC8, which is described in more detail in Appendix 6. This vector has been used to clone duplex cDNA molecules by the double-linker approach (Kurtz & Nicodemus 1981, Helfman *et al.* 1983), in which *different* linker molecules are added to the opposite ends of the cDNA (Fig. 2.8). This has the following advantages.

1 The problem of vector reclosure without insertion of foreign DNA is overcome. Partly for this reason the method is efficient, i.e. cDNAs have been cloned which were derived from rare mRNA molecules in the starting population.

2 The use of linkers, rather than homopolymers, is desirable when expression from a vector-borne promoter is sought (see Chapter 7).

3 The orientation of the inserted DNA is fixed.

Adaptors

It may be the case that the restriction enzyme used to generate the cohesive ends in the linker will also cut the foreign DNA at internal sites. In this situation the foreign DNA will be cloned as two or more subfragments. One solution to this problem is to choose another restriction enzyme, but there may not be a suitable choice if the foreign DNA is large and has sites for several restriction enzymes. Another solution is to methylate internal restriction sites with the appropriate modification methylase. An example of this is described in Chapter 5. Alternatively, a general solution to the problem is provided by chemically synthesized adaptor molecules which have a *preformed* cohesive end (Wu *et al.* 1978). Consider a blunt-ended foreign DNA containing an internal *Bam* HI site (Fig. 2.9), which is to be cloned in a *Bam* HI-cut vector. The *Bam* adaptor molecule has one blunt end, bearing a 5'-phosphate group, and a Bam cohesive end which is not phosphorylated. The adaptor can be ligated to the foreign DNA ends, without any risk of adaptor self-polymerization. The foreign DNA plus added adaptors is then phosphorylated at the 5'-termini and ligated into the *Bam* HI site of the vector. If the foreign DNA were to be recovered from the recombinant with *Bam* HI, it would be obtained in two fragments. However, the adaptor is designed to contain two other restriction sites (*Sma* I, *Hpa* II) which may enable the foreign DNA to be recovered intact.

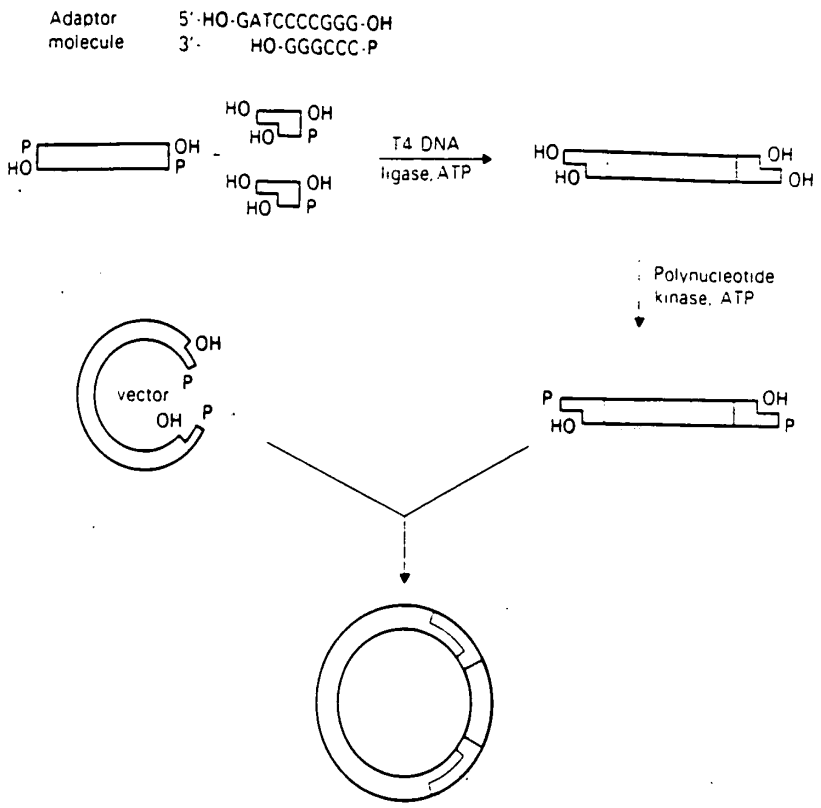


Fig. 2.9 Use of a *Bam* HI adaptor molecule. A synthetic adaptor molecule is ligated to the foreign DNA. The adaptor is used in the 5'-hydroxyl form to prevent self-polymerization. The foreign DNA plus ligated adaptors is phosphorylated at the 5'-termini and ligated into the vector previously cut with *Bam* HI.

Homopolymer tailing

A general method for joining DNA molecules makes use of the annealing of complementary homopolymer sequences. Thus, by adding oligo (dA) sequences to the 3' ends of one population of DNA molecules, and oligo (dT) blocks to the 3' ends of another population, the two types of molecule can anneal to form mixed dimeric circles (Fig. 2.10).

An enzyme purified from calf-thymus, terminal deoxynucleotidyl-transferase, provides the means by which the homopolymeric extensions can be synthesized, for if presented with a single deoxynucleotide triphosphate it will repeatedly add nucleotides to the 3'-OH termini of a population of DNA

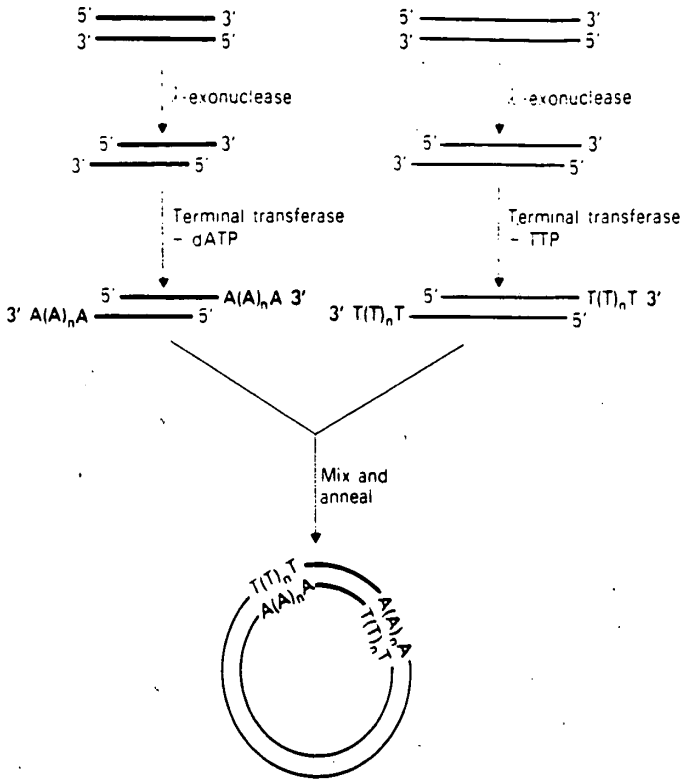
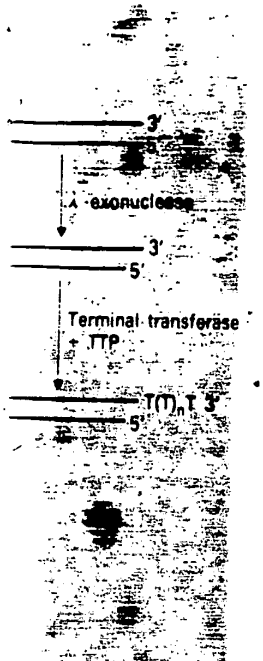


Fig. 2.10 Use of calf-thymus terminal deoxynucleotidyl-transferase to add complementary homopolymer tails to two DNA molecules.

molecules (Chang & Bollum 1971). DNA with exposed 3'-OH groups, such as arise from pretreatment with phage λ exonuclease or restriction with an enzyme such as *Pst* I, is a very good substrate for the transferase. However, conditions have been found in which the enzyme will extend even the shielded 3'-OH of 5'-cohesive termini generated by *Eco* RI (Roychoudhury *et al.* 1976, Humphries *et al.* 1978).

The terminal transferase reactions have been recently characterized in detail with regard to their use in gene manipulation (Deng & Wu 1981, Michelson & Orkin 1982). Typically, 10–40 homopolymeric residues are added to each end.

In 1972, Jackson *et al.* were among the first to apply the homopolymer method when they constructed a recombinant in which a fragment of phage λ DNA was inserted into SV40 DNA. In their experiments, the single-stranded gaps which remained in the two strands at each join were repaired *in vitro* with DNA



ıcleotidyl-transferase to
DNA molecules.

with exposed 3'-OH
with phage λ exo-
ıch as *Pst* I, is a very
ver, conditions have
nd even the shielded
y *Eco* RI (Roychoud-

been recently charac-
se in gene manipu-
kin 1982). Typically
each end.
ıe first to apply the
ted a recombinant in
rted into SV40 DNA.
aps which remained
d *in vitro* with DNA

polymerase and DNA ligase so as to produce covalently closed circular molecules, which were then used to transfect susceptible mammalian cells (see Chapter 11).

Subsequently, the homopolymer method, employing either dA.dT or dG.dC homopolymers, has been applied extensively in constructing recombinant plasmids for cloning in *E. coli*. Commonly, the annealed circles are used directly for transformation with repair of the gaps occurring *in vivo*. In the example which follows we shall see how homopolymer tailing can be applied to cloning DNA copies of eukaryotic messenger RNA and how a careful choice of which homopolymers are used can be important.

Cloning cDNA by homopolymer tailing

If we wish to construct a clone containing sequences derived from eukaryotic mRNA, we must first obtain the sequence in DNA form. We can do this by making a complementary (cDNA) copy of the mRNA, using the enzyme reverse transcriptase, which is a type of DNA polymerase found in retroviruses, and whose function is to synthesize DNA upon an RNA template.

Like other true DNA polymerases, reverse transcriptase can only synthesize a new DNA strand if provided with a growing point in the form of a pre-existing primer which is base-paired with the template and bears a free 3'-OH group. Fortunately, most eukaryotic mRNAs occur naturally in a polyadenylated form with up to 200 adenylate residues at their 3'-termini and hence we can provide a primer simply by hybridizing a short oligo (dT) molecule with this poly (A) sequence. The primer is then suitably located for synthesis of a complete cDNA by reverse transcriptase in the presence of all four deoxynucleoside triphosphates (Fig. 2.11).

The immediate product of the reaction is an RNA–DNA hybrid. The RNA strand can then be destroyed by alkaline hydrolysis, to which DNA is resistant, leaving a single-stranded cDNA which can be converted into the double-stranded form in a second DNA polymerase reaction. This reaction depends upon the observation that cDNAs can form a transient self-priming structure in which a hairpin loop at the 3'-terminus is stabilized by enough base-pairing to allow initiation of second-strand synthesis. Once initiated, subsequent synthesis of the second strand stabilizes the hairpin (Efstratiadis *et al.* 1976, Higuchi *et al.* 1976). The hairpin and any single-stranded DNA at the other end of the cDNA molecule are then trimmed away by treatment with the single-

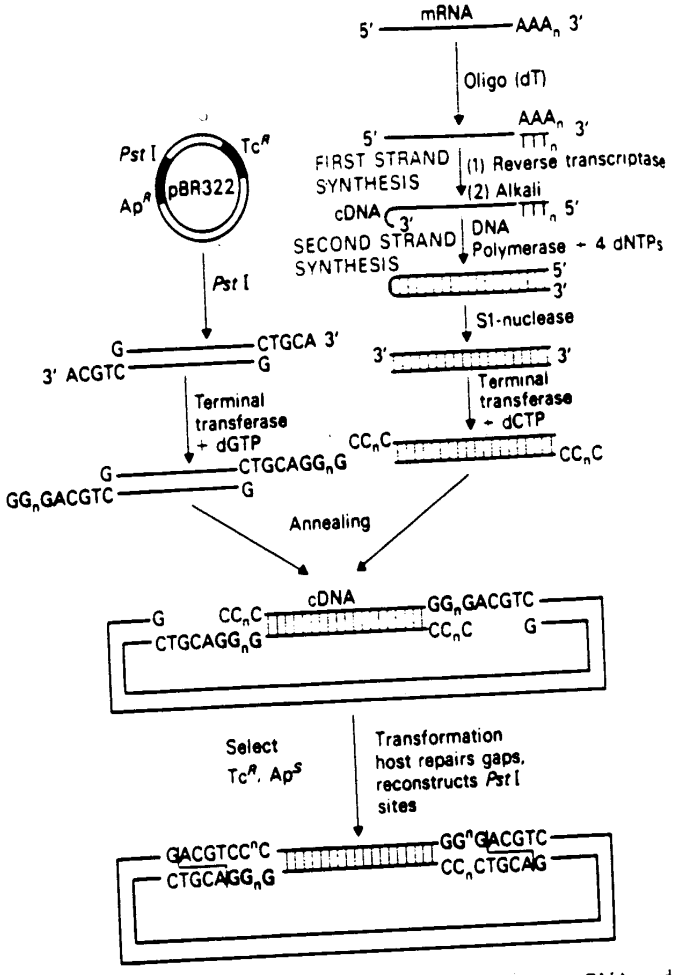


Fig. 2.11 Synthesis of a cDNA copy of a polyadenylated mRNA and insertion into a vector molecule by homopolymer tailing. See text for explanation.

strand-specific nuclease S1, giving rise to a fully duplex molecule.

In our example (Fig. 2.11) the duplex cDNA is tailed with oligo (dC) and annealed with the pBR322 vector which has been cut open with *Pst* I and tailed with oligo (dG). It will be seen that these homopolymers have been chosen so that *Pst* I target sites are reconstructed in the recombinant molecule, thus providing a simple means for excising the inserted sequences after amplification (Smith *et al* 1979) This can be accomplished in another way, by constructing dA dT joins In that case the

homopolymeric regions will have a lower melting temperature than the rest of the recombinant molecule, and so under partially denaturing conditions can be cleaved by nuclease S1 to release the inserted sequence (see p. 247 for example).

Full-length cDNA cloning

In the previous cDNA cloning scheme second strand synthesis is self-primed, resulting in the formation of a duplex cDNA with a hairpin loop that is subsequently removed by nuclease S1. This step necessarily leads to the loss of a certain amount of sequence corresponding to the 5′ end of the mRNA, and unless the nuclease S1 is very pure, there can be adventitious damage to the duplex cDNA.

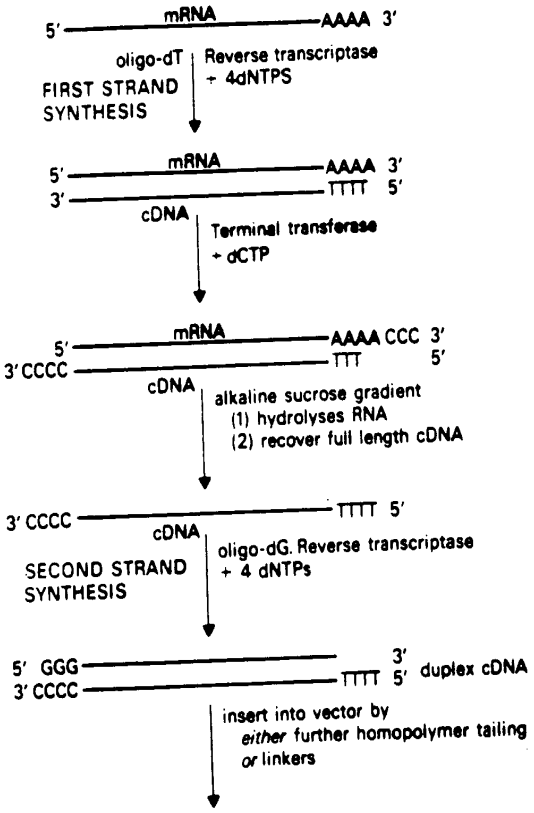
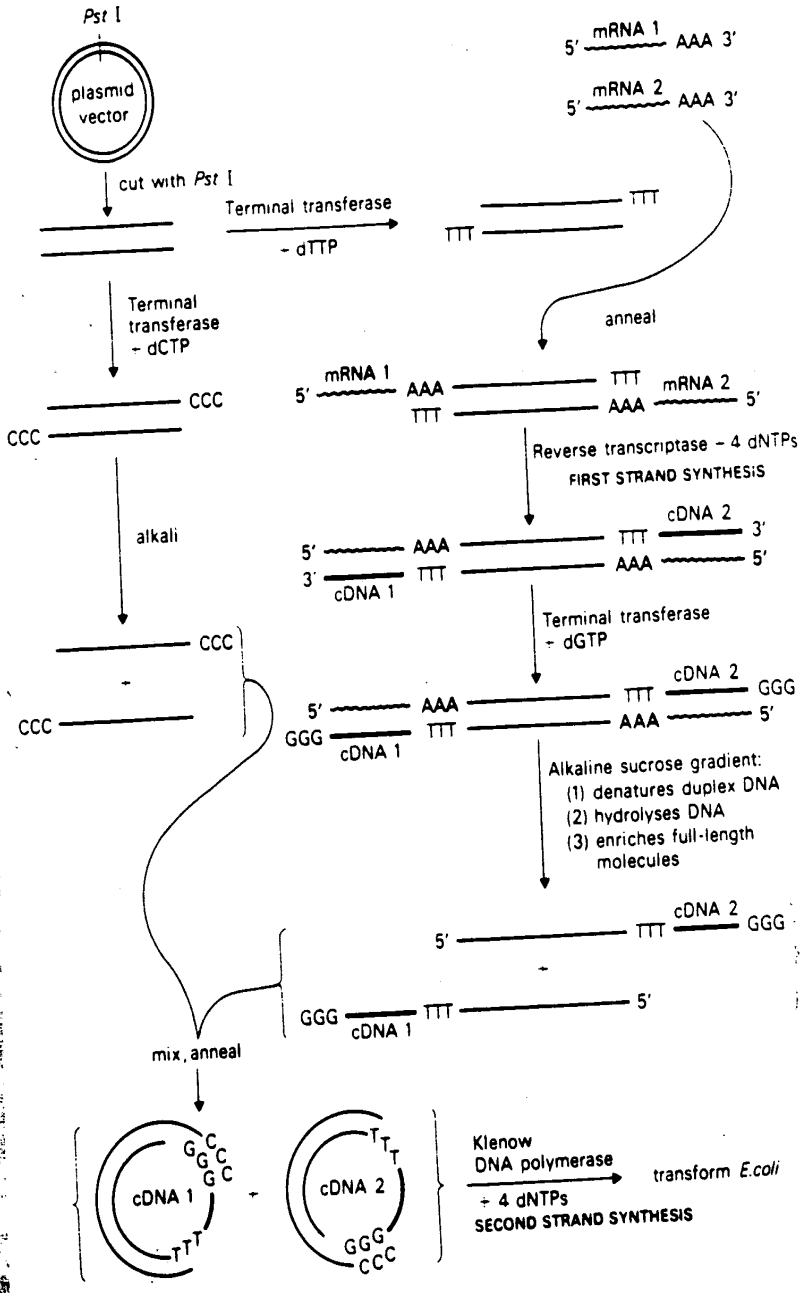
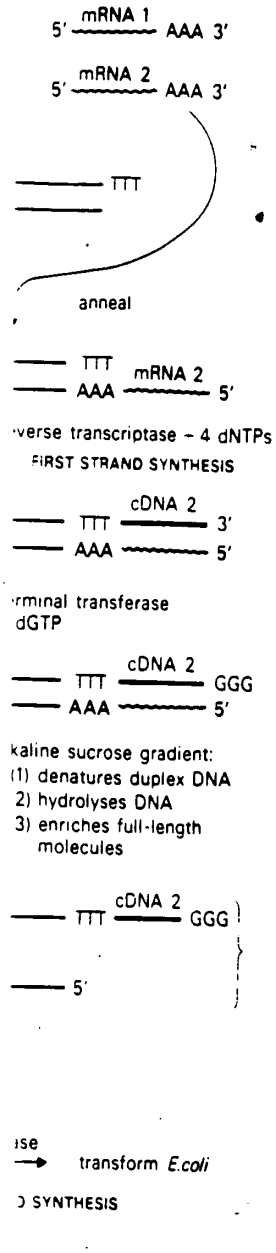


Fig. 2.12 Improved method for full-length duplex cDNA synthesis. The first strand is tailed with oligo(dC) so as to allow priming of the second strand synthesis by oligo (dG).


Pst I
plasmid vector
cut with Pst I
Terminal transferase
+ dTTP
Terminal transferase
+ dCTP
alkali
mix, anneal
mRNA 1
mRNA 2
anneal
Reverse transcriptase + 4 dNTPs
FIRST STRAND SYNTHESIS
cDNA 1
cDNA 2
Terminal transferase
+ dGTP
Alkaline sucrose gradient:
(1) denatures duplex DNA
(2) hydrolyses DNA
(3) enriches full-length molecules
Klenow
DNA polymerase
+ 4 dNTPs
SECOND STRAND SYNTHESIS
transform E.coli



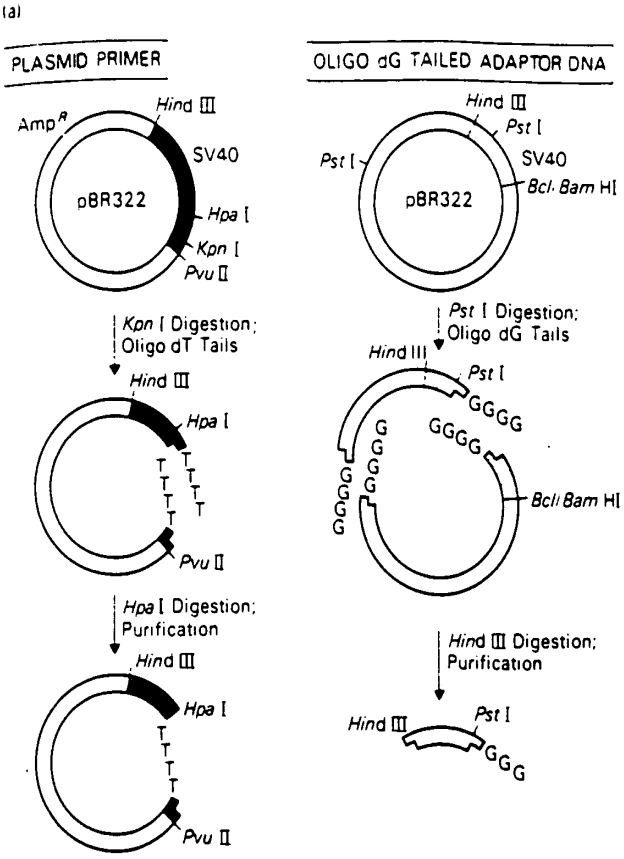


Fig. 2.14 The Okayama and Berg method of cDNA cloning. (a) Preparation of plasmid primer and adaptor DNA. The unshaded portion of each ring is pBR322 DNA, and the shaded or stippled segments are from SV40 DNA.

Fig. 2.13 (*facing page*) Efficient full-length cDNA cloning (Heidecker & Messing 1983). The mRNA is annealed to linearized and oligo-dT tailed plasmid DNA, which then primes synthesis of the first cDNA strand. Oligo-dG tails are added to the cDNA-plasmid molecules, which are then centrifuged through an alkaline sucrose gradient. This step removes small molecules, hydrolyses the mRNA and separates the two cDNAs which were formerly attached to the same duplex plasmid. Denatured, oligo-dC tailed plasmid DNA is added (in excess) and conditions adjusted to favour circularization by the complementary homopolymer tails. The excess oligo-dC tailed plasmid may simply renature, but cannot circularize. The circular molecules have a free 3'-hydroxyl on the oligo-dC tail which primes second strand synthesis of the cDNA to create duplex recombinant plasmids which transform *E. coli*. Clones can be obtained with the cDNA inserted in both orientations.

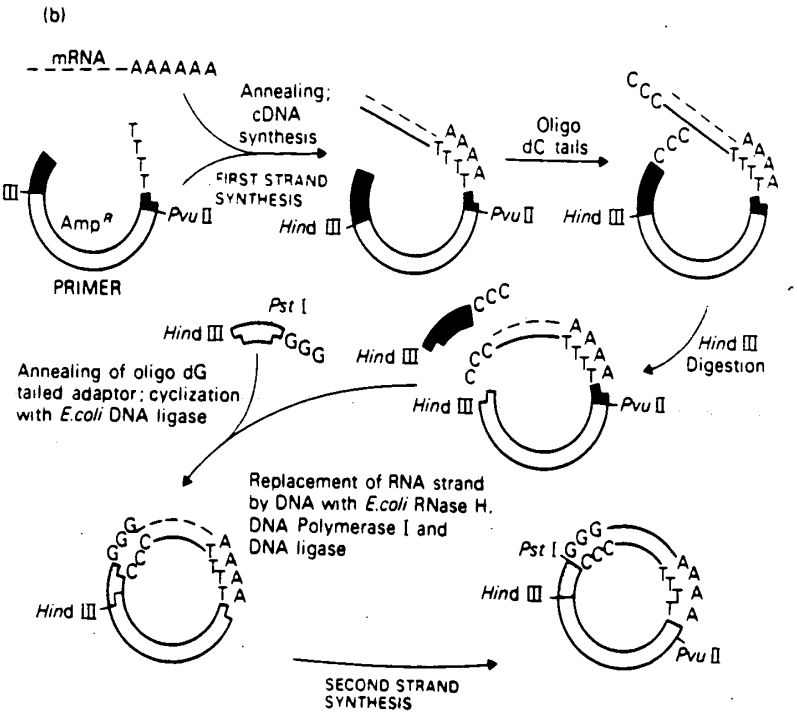


Fig. 2.14 (b) Steps in the construction of plasmid-cDNA recombinants. The designations for the DNA segments are as mentioned in (a).

dC tailing of single-stranded cDNA followed by oligo-dG priming of second strand synthesis (Fig. 2.12) does not lead to hairpin formation, nuclease S1 treatment is not required, and consequently is an effective method for generating full-length cDNA clones (Land *et al.* 1981). Two further methods, shown schematically in Figs 2.13 and 2.14, have been devised to eliminate the use of nuclease S1. Additionally, in both methods the oligo-dT sequence for priming the first strand cDNA synthesis is linked to the vector DNA in a prior reaction. Both methods have been reported to promote full length cDNA cloning with a very high efficiency (Okayama & Berg 1982, Heidecker & Messing 1983). It is thought that full length reverse transcripts are obtained *preferentially* because in each case an RNA–DNA hybrid molecule, which is the result of first strand synthesis, is the substrate for a terminal transferase reaction. A cDNA that does not extend to the end of the mRNA will present a shielded 3'-hydroxy group, which is a poor substrate for tailing.